International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
***Article Received:** 25 January 2023 **Revised:** 12 February 2023 **Accepted:** 30 March 2023*

Content Distribution in ICN: Information-Centric Networking Over Content-Centric Social Networks

Prashanth Kumar P[1] , Dr. Pramod Pandurang Jadhav[2]
[1]Research scholar
[1,2]Department of Computer Science & Engineering
[1,2]Dr. A.P.J. Abdul Kalam University, Indore (M.P.) - 452010

Abstract : The proliferation of digital gadgets and the popularity of social networking have led to many people documenting and broadcasting their private lives online. Content sharing has become one of the most common online activities, yet it is rarely considered during the design phase of a website. Information-centric networking (ICN) is a novel architecture developed to handle the rapidly evolving Internet usage patterns. ICN optimizes information exchange by treating content as the fundamental component of networking functions. This study presents a method that enhances ICN's core functionality by leveraging content identification at the network layer. This enables the selective retrieval of material parts already stored in ICN's in-network caches, allowing information producers to offload some content processing tasks to the network. The findings show reduced network core traffic and storage usage. To address content push issues, precise user descriptions and support for the big data nature of both individuals and content are necessary. Content-centric networking (CCN) offers a solution with lower connection costs and faster download times. By using named content, CCN facilitates real-time interest tracking, enabling the study of new content and enhancing online learning algorithms. This approach results in efficient outcomes in the big data context, avoiding additional network strain and overcoming the cold start problem.

Keywords: *MHR (maximum heat ratio), collaborative learning, big data, social networks*

I. INTRODUCTION

Interest in Information-Centric Networking (ICN/CCN) has grown in recent years, mostly as a result of the fact that the fundamental purpose of the Internet is now the delivery and consumption of information morsels. Several research centres have suggested and developed [1] content-centric architecture. Historically, hosts have been at the centre of the Internet's operations, creating packets that are then routed to other servers. For the better part of the Internet's history, this function has been essential to its operation, driven by the needs of widely used higher-level applications like "hypertext transfer protocol (HTTP), electronic mail (e-mail), and file transfer protocol (FTP)" [2, 3].

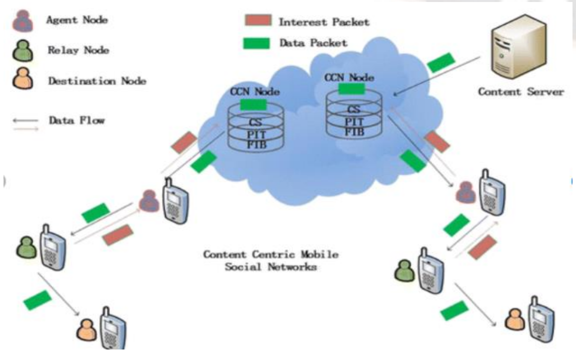

Figure 1: Schematic Representation of Content-Based Social Networks

Establishing an end-to-end connection with them is now the only viable option for obtaining stuff. Through its history, the Internet has undergone a number of significant changes to its communication pattern, from the client-server model to peer-to-peer networking and finally to cloud networking. However, one thing has remained constant: the Internet's usage pattern has become increasingly content-oriented [5]. In other words, consumers of material don't care how or where it was obtained. However, the internet wasn't meant to be a media delivery network; it was built to facilitate communication. When it comes to wired and wireless communications, these constraints affect every player in the ecosystem, thus operators require cost-effective strategies to overcome them [6].

Nowadays, people primarily care about obtaining material and using services, regardless of where those things are offered. In order to take advantage of these use situations, new networking technologies have been developing.

There are a variety of ICN implementations follow different conceptual concepts (such as communication models as well as the naming conventions, routing as well as forwarding). This study focuses upon Named Data Networking (NDN) protocols that are used for interest-based ICN solutions. The protocol is designed to accommodate users with an interest in accessing particular types of information. The title could be considered the data selection mechanism in the networks

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
***Article Received:** 25 January 2023 **Revised:** 12 February 2023 **Accepted:** 30 March 2023*

in which data is distributed. The ability to use network cache is their major benefit. The granular caching can only be achieved if that the creator of content provides information that is of the appropriate resolution. The benefits of ICN caching, therefore, are not realized even though content (dis)aggregation could significantly meet the requirements of the requirements of data centres. In addition, delegating the task of provisioning data is the desired outcome in distributing content across the internet, however there's no way to do this to enable in-network content (dis)aggregation.

Aspects to consider when implementing every network system should comprise considerations regarding security of information as well as privacy. Controlling access, confidentiality authenticity, integrity, and accessibility prerequisites are crucial to encourage adoption across the board of any system.

When taking into account circumstances like the data centre, which closely connect the physical and digital worlds, this becomes more pertinent. Despite the fact that ICN incorporates security into its architecture, a sizable number of criteria still need to be solved in order to maintain the Data centre safely.

Additionally, there are security issues that arise from the implementation of functions in the network that are required for ICN to be considered with regard to the particular circumstances of the current proposed proposal. Our work has two significant security flaws that are linked to trust and access control. Cached data may be processed by intermediary nodes to create "new" data. Since the information's signer won't also be its creator, the original ICN signature verification method is broken. We foresee approaches to this problem that depend on schematized trust and the use of local trust anchors. Access control techniques are also put to the test by the potential for selective access to certain bits of material. In our hypothetical scenario although a particular user would not be granted access to the entire name-based access control. However, these security-related procedures link to broader issues with the NDN operation as a whole, therefore it should be the subject of more intensive activities that are specifically targeted at those elements. Because of this, the appropriate implementation and design of security methods are beyond the purview of this work.

CCN is a different networking architecture that is built on the idea that a user should be able to concentrate on the data they need without worrying about where it is physically stored. CCN offers content caching, which decreases network congestion and speeds up delivery, simplifies the deployment of network devices, and incorporates data-level security into the network itself [7]. With the publication of the open source code and the Android implementation, as well as the commercial partnerships with notable industrial partners, the programme has continued to gather pace.

CCN differs from the standard TCP/IP protocol stack in the ways described below:

- In the pull model of communication, the receiver initiates contact by signalling their interest in a topic. A single data message is sent in response to an interest.
- The Content Naming Convention (CCN) is a hierarchical naming method for content, rather than an address for individual hosts. Like URLs, content is given hierarchical names. The longest-prefix matching algorithm is used to determine which packets of interest to forward.
- Cache and forward architecture: Data may be cached and used to serve subsequent requests from any CCN device.

Big Data

Big Data is defined as data collections so large that they cannot be effectively stored, processed, or analysed using traditional methods.

Data is being produced at a dizzying pace from the millions of data sources available today. All around the globe, there are these data sources. Social media platforms and networks are among the most significant data sources [8]. To give you an idea, Facebook generates over 500 GB of data every single day. This information may take a variety of forms, including pictures, videos, texts, and more. The different types of data include both organised and unstructured varieties [9]. Big Data refers to this accumulated data. Big Data, however, is useless in its unprocessed state. Now that we've covered Big Data Analytics, let's talk about it. Consider these four benefits of big data analysis.

Big Data Analytics: Pros and Cons

1. Management of Risk

Use case: Using Big Data analytics, Banco de Oro, a Philippine banking institution, can spot irregularities and fraud. The company utilises it to determine which factors are really at play when issues arise.

2. Product Creation and Innovation

Use Case: One of the major producers of jet engines for airlines and military forces throughout the world, Rolls-Royce, using Big Data analytics to evaluate the efficacy of its engine designs and identify areas for development.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
***Article Received:** 25 January 2023 **Revised:** 12 February 2023 **Accepted:** 30 March 2023*

3. Organizational Decision Making that is Faster and Better

Use Case: To aid in its decision-making process, Starbucks employs Big Data analytics. The corporation uses it to determine, for instance, whether a certain area is appropriate for the opening of a new store. A number of aspects, including population, demography, accessibility, and others, will be examined.

4. Improve Customer Experience

Use Case: In order to better serve its customers, Delta Air Lines analyses big data. They keep an eye on Twitter to see what consumers are saying about their trips, delays, and so on. When the airline sees a tweet with a poor review, it takes action. The airline may improve its standing with customers by openly addressing their concerns and providing workable answers.

Big Data Analytics' Lifecycle Phases

Let's examine Big Data analytics' fundamentals:

- The first phase in the Big Data analytics lifecycle is the business case, which describes the justification and goals of the research.
- In the second phase, known as "data identification," several data sources are discovered.
- In Stage 3, we filter all the data that was detected in Stage 2 to get rid of any bad data.
- At this point, data that is incompatible with the tool is extracted and changed into a format that is.
- Data with the same fields from several data sources are combined in Stage 5 (Data Aggregation).
- The sixth stage, data analysis, involves evaluating the data using analytical and statistical methods to extract insights.
- The seventh step is data visualisation, where Big Data analysts may use tools like Tableau, Power BI, and QlikView to graphically portray their results.
- In Stage 8 of the Big Data analytics lifecycle, the final analysis results are made available to business stakeholders, who will then take action.

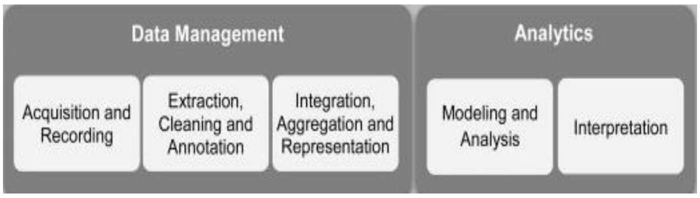

Figure 2: Big Data Processes

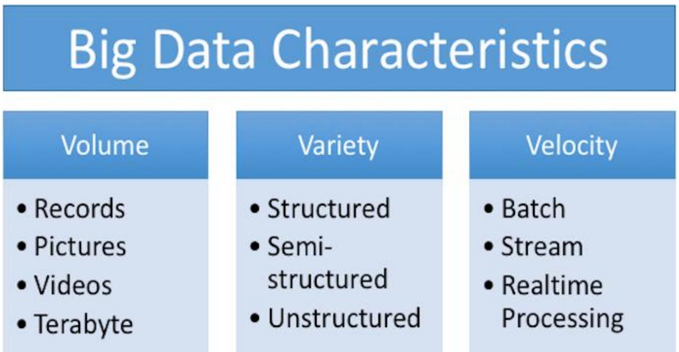

Figure 3: Big Data Online Learning System

II. RESEARCH METHODOLOGY

Algorithm based on collaborative training

The heat ratio algorithm is based on the collaborative training explanation detailed. The algorithm tries to optimize the temperature of a device using data that has been condensed in a certain way. The algorithm is composed of two parts: the first part uses the values of the temperature variable to learn how much the device needs to be heated in order to reach a certain value, while the second part tries to find the best values of theTemperature variable for each image to find the most efficient device. User post ratings provide an initial score matrix for the collaborative training-based hybrid recommendation system. To calculate the composite score, a viewpoint pre-filtering scoring matrix is generated. Fig. 3 shows a cooperative training-based hybrid recommendation system. Rm n(U, P) is the scoring matrix for m users and n posts. User u's score on post p is Ru(p), where p is the HITS authority value.

The user's actual scoring matrix Rm n(U, P) and the emotion analysis model's simulated scoring matrix Rmn(U, P) serve as inputs and outputs, respectively, for the object-based collaborative filtering recommendation model on dataset D train. Posting-based collaborative filtering is described in Algorithm 1.

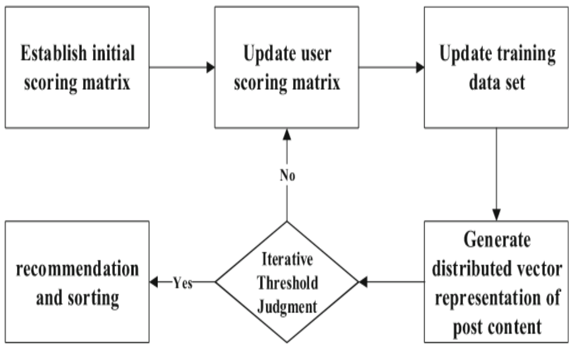

Figure 4: Block diagram: MHR Algorithm based on collaborative training

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
***Article Received:** 25 January 2023 **Revised:** 12 February 2023 **Accepted:** 30 March 2023*

Algorithm: MHR

1. Calculate a probabilistic cumulative distribution function (pdf) for the items in the dataset.

2. Divide the dataset into k subsets, where k is the number of items.

3. For each subset, calculate a heat ratio for every item in the subset.

4. Aggregate the heat ratios across all subsets to produce a new dataset.

5. Repeat steps 2-4 until the data is exhausted.

User u's training data set and score matrix default value are modified using post-based collaborative filtering recommendation. Emotional classification often uses fine-grained and coarse-grained differences. The recommendation system in this research uses 2-level emotional categorization since it is more accurate than the 5-level model. The emotion metre employed 1 and 0 to express the user's sentiments.

Data selection in collaborative training

The collaborative training model is built with the addition of a data selection approach to choose which data should be included in the training set. The data is divided up into categories that correspond with the user's level of experience. Data in the training set is labelled, whereas prediction data is unlabeled. The samples are randomly distributed within each cluster, and the confidence score of the sample falls into one of many predetermined categories in the data selection approach.

Simulation Set-Up:

A binary tree structure with changeable depth is included in the simulated scenario. The information producer is the node at the base of the tree, followed by NDN routers and information consumers in the subsequent levels. The CiC forwarding technique is offered at the edge layer of routers, which is the layer closest to the consumers. In the course of 20 seconds during the study, users communicate Interests at a speed of 10 interest per second. For each interest, the question which is requested (i.e. the one from the below: (i) speed; (ii) direction; (iii) wheel_count; (iv) the entire contents) will be randomly selected. According to Figure 3. The material generated by the producer is content that has been encoded with JSON. Three freshness levels are considered (i) the value of 0 seconds is a sign that the cached content isn't suitable; (ii) 50 s means that the data is stored at once and suitable throughout the simulation (iii) 10 sec and calls for it to refresh the cache every few seconds every 10 seconds throughout the simulation. The primary variables utilized to design the simulation can be found in Table 1. An identical scenario using the identical operating configuration but with no CiC upgrade was also replicated to create the basis for comparison.

Table 1. Simulation set-up parameters.

Parameter	Value
Channel_Data_Rate	10_Mbps
Latency	100_ms
Interest_Rate	100_Interests/s
CS_Size	110_packets
CS_Replacement_Policy	LRU
Freshness	[0,_20,_100]_s
Topology_depth	(4,_16)
Simulation_Duration	40_s

III. RESULT ANALYSIS

Model N's iterations for classifying emotions

We employ an embedding approach based on user ratings and a pre-filtering strategy based on views to merge user ratings and emotional prediction ratings, respectively, in order to account for the interplay between these two sets of data.

The simulation findings are discussed after being given. As was previously described, three distinct metrics were assessed, each of which addressed a different research issue and was computed for a particular topological depth and data freshness value. The second is the average hop count between data packets, which is determined as the total number of data packets divided by the total number of hops for each data packet that consumers have received.

MHR

To bridge the gap between collaborative filtering recommendation and content-based recommendation for articles, we provide a new mechanism we term MHR (maximum heat ratio). Both the subjective nature of user ratings and the objective meaning of a post's natural language description are taken into account when calculating the recommendation factor. In addition, the jumbled data collected from social networks is processed using the data pre processing approach, which eliminates the majority of the poor quality individuals and posts. This supporting data should, in theory, make it easier to deal with cold starts and limited data.

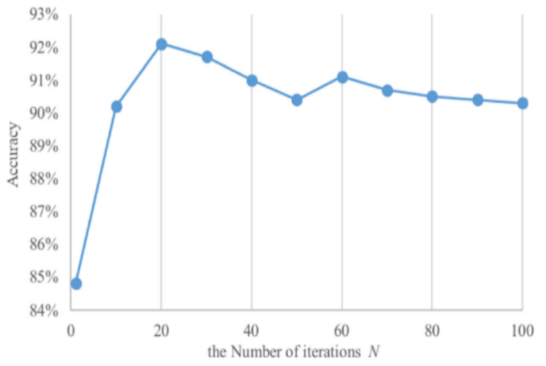

Figure 5: Iteration of classification of Model N

The experimental findings show that both the suggested MHR method and the HFT algorithm outperform their competitors on N of the evaluation indices. These two algorithms are more effective at dealing with the recommendation system's cold start issue because they mine user comment data and utilise post content description text data in the collaborative training model.

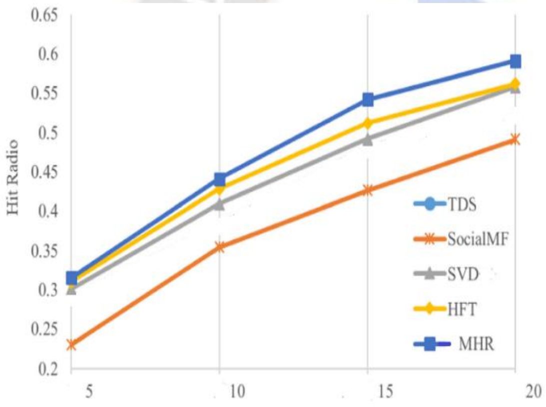

Figure 6: Hit Radio Comparison

IV. CONCLUSION

We can accurately profile users with the help of CCN, allowing us to send them tailored material. We provide a contextual method to online learning that works with the continual inflow of large data to generate accurate pushes in CCN-based social networks. In order to increase the efficiency of CCN, our method aids in decreasing the amount of duplicate data transformation in existing networks. Since our MHR method only adds a little amount of extra work for service providers, but greatly improves push performance, we show that it achieves sub-linear regret and space complexity. We have found that our method is an excellent fit for this use case. We want to evaluate the efficacy of our push mechanism in the future by putting it to use in a functional CCN-based social network. We'd also want to make our algorithm more efficient with regards to energy use.

Reference

[1] E. Zeydan et al., "Big Data Caching for Networking: Moving from Cloud to Edge," IEEE Communication. Mag., vol. 54, no. 9, Sept. 2016, pp. 36–42.

[2] Z. Chang et al., "Collaborative Mobile Clouds: An Energy Efficient Paradigm for Content Sharing," IEEE Wireless Communication, 2017.

[3] M. S. ElBamby et al., "Content-Aware User Clustering and Caching in Wireless Small Cell Networks," Proc. 11th Int'l.Symp. Wireless Communication. Systems, Barcelona, Spain, Aug.2014.

[4] Z. Han, M. Hong, and D. Wang, Signal Processing and Networking for Big Data Applications, Cambridge Univ. Press.

[5] K. Guo, C. Yang, and T. Liu, "Caching in Base Station with Recommendation via Q-Learning," 2017 IEEE Wireless Communication. Networking Conf., San Francisco, CA, Mar. 2017.

[6] J. Song et al., "Learning Based Content Caching and Sharing for Wireless Networks," IEEE Trans. Communication., 2017.

[7] B. N. Bharath, K. G. Nagananda, and H. V. Poor, "A Learning-Based Approach to Caching in Heterogenous Small Cell Networks," IEEE Trans. Commun., vol. 64, no. 4, Apr. 2016

[8] L. Lei et al., "A Deep Learning Approach for Optimizing Content Delivering in Cache-Enabled HetNet," Proc. IEEE Int'l. Symp. Wireless Communication. Systems, Bologna, Italy, Aug. 2017.

[9] C. B. Browne et al., "A survey of Monte Carlo tree search methods," IEEE Trans. Comput. Intell. AI Games, vol. 4, no. 1, pp. 1–43, Mar. 2012.

[10] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," Mach. Learn., vol. 47, pp. 235–256, 2002.

[11] J. Li, B. Liu, and H. Wu, "Energy-efficient in-network caching for content-centric networking," IEEE Communication. Lett., vol. 17, no. 4, pp. 797–800,Apr. 2013.

[12] A. Gharaibeh, A. Khreishah, and I. Khalil, "An O(1)-competitive online caching algorithm for content centric networking," in Proc. 35th Annu. IEEE Int. Conf. Computer. Communication, 2016, pp. 1–9.

[13] Y. Kim and I. Yeom, "Performance analysis of in-network caching for content-centric networking,"

Computer. Network, vol. 57, no. 13, pp. 24652482, 2013.

[14] J. Quevedo, D. Corujo, and R. Aguiar, "Consumer driven information freshness approach for content centric networking," in Proc. IEEE Conf. Computer. Communication. Workshops, Toronto, ON, Canada, 2014, pp. 482–487

[15]C. Jiang et al., "Machine Learning Paradigms for Next-Generation Wireless Networks," IEEE Wireless Commun., vol. 24, no. 2, Apr. 2017, pp. 98–105.